FACTORIAL || JOURNAL OF ENGINEERING, MANAGEMENT, AND TECHNOLOGY
Volume-I || Issue-I || July-Dec, 2025

Exploring Machine Learning Methods: A Detailed Survey

Amritpreet Kaur[1], Komal Sharma[2]
[1,2]*Assistant Professor, Dept. of Computer Applications, Sri Aurobindo College of Commerce and Management, Ludhiana*
amritpreet@saccm.in, komalsharma00061@gmail.com

ABSTRACT

Machine learning (ML) is a dynamic area of computer science that enables systems to learn and make decisions without being explicitly programmed. It is particularly useful in solving complex computational problems where designing rule-based algorithms is impractical. ML has found widespread applications across domains such as artificial intelligence, data analytics, and pattern recognition. However, developing efficient ML applications requires a solid understanding of intelligent systems, algorithms, and their underlying mechanisms. This survey presents a comprehensive overview of machine learning, focusing on its core principles, commonly used techniques, and algorithmic foundations. It examines various ML approaches—including supervised, unsupervised, and reinforcement learning and analyzes how their effectiveness varies across different application domains. Additionally, the paper discusses the challenges involved in deploying ML systems, including issues related to performance optimization, data infusion, and real-world impact. A key contribution of this survey lies in its comparative analysis of ML techniques across diverse use cases, offering insights into their strengths, limitations, and suitability for specific problem types. By synthesizing existing knowledge and highlighting current trends, this paper aims to serve as a valuable resource for both researchers and practitioners seeking to understand and apply machine learning effectively.

Keywords: *Machine Learning, Supervised Learning, Semi-Supervised Learning, Unsupervised Learning, Artificial Intelligence, Artificial Neural Network*

ARTICLE INFORMATION

Available online: 22 September 2025

DOI

10.5281/zenodo.17239950

1. Introduction

Machine Learning (ML) can be regarded as a subfield of Artificial Intelligence (AI), as these algorithms can be viewed as constructing blocks to allow computers to apprehend the way to continue significantly even more shrewdly, rather than simply putting things away and retrieving information like a database framework and various applications. It enables frameworks to gain from their experience and predict more accurate outcomes based on that experience. Machine learning is driven by several scholastic orders, including informatics, measurements, science and brain study. Machine learning's center target is to enhance computers because they normally find a not too bad marker based on past experiences, and an exceptional classifier performs this mission.

Collecting is the best way to use a model to predict unknown characteristics (yield factors), using commonly known characteristics (input factors). It can be used in various speculative disciplines, such as Neuroscience, Statistics, Psychology, Astronomy, etc. [1].

ML algorithms can also be effectively implemented to effectively classify text. Classification is the method to predict an unknown value using predefined values with the help of a model. The fundamental procedure of ML is to provide the learning algorithm with training data. It generates another set of classification rules based on data interpretations. The new generated machine learning algorithm can be applied to different categories of data and thus, helps to generate new models. For example, a similar type of learning algorithm could be used for stock prediction as well as language translation. The main power of ML is to deduce new instructions from a huge amount of social media data. In most of the machine learning methods, the apprentice is not aware of the action being taken but instead has to determine which actions attain the maximum accuracy by implementing. Figure 1 shows the general structure building [2].

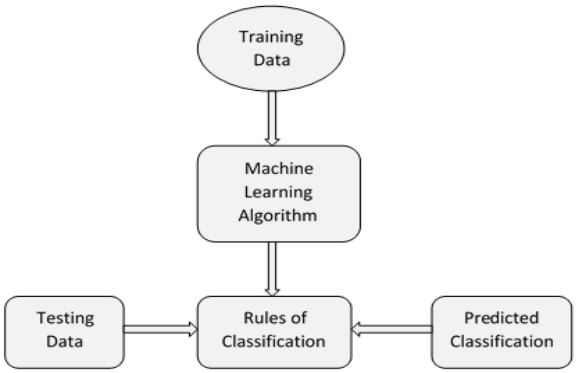

Fig.1. Classification Architecture

The beginning stage of the fundamental model represented by Figure 2 illustrates that during the time spent in machine learning, the nature of data that the outer condition gives to the framework is the essential factor. The outer condition is outside data set that conveys itself in some structure, it speaks to wellsprings of outside data; Learning is the procedure that procedures the outside data to knowledge, first it gets the data of outside condition and afterward forms the data to knowledge, and places these knowledge into the archive; Repository stores many general rules that guide a piece of the execution activity, because of condition gives a wide range of data to learning framework, the nature of data impacts legitimately on learning acknowledgment whether simple or tumultuous. The second factor that affects the plan of learning framework is the storehouse. The declaration of knowledge is shifted, for example, eigenvector, logic explanations of the principal request, generation rules, semantic networks and structures, and so forth; each style of articulation has its solid point [3].

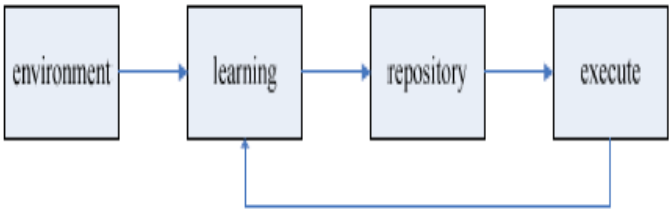

Fig. 2. The basic model of Machine Learning

I. LITERATURE SURVEY

Raouf Boutaba et.al. [4] gives a systematic analysis of various applications of ML techniques related to networking technologies. According to the authors, machine learning is capable of solving diverse problems due to data accessibility, improved ML techniques and computational capabilities. Danijel Kučak et.al. [5] explains that ML is one of the most encouraging approaches used in the field of Information Technology. The utilization of ML in the field of education is as of now intriguing to specialists and researchers. The paper focuses to gauge the current patterns and the potential outcomes of applying and utilizing ML in the education sector.

S. Sarojini Devi et.al. [6] highlights the importance of the Bankruptcy Prediction Model (BPM) that is distinctively fundamental for money related foundations to check the financial soundness of organizations or the executives. This paper presents a detailed study of various ML techniques to handle bankruptcy issues and the future study includes the improved approach by using the Apache Mahout tool. T. Sree Kala [7] In the earlier decades, the quick advancement of Intrusion Detection and Prevention frameworks assumed a vital job in computer networks and security. This paper discusses various ML methods to ensure cybersecurity and detect malicious data access. The Intrusion Detection System (IDS) can handle the data over the network independently without any human involvement. But it is difficult for a human to identify the intrusions in network traffic.

Astina Minz [8] proposed a model in which the picture preparation and machine learning systems were utilized for the discovery of the tumor. Median channels are utilized for cleaning the pictures', thresholding method is utilized for the division of the mind, GLCM procedure is utilized for extricating the highlights, and lastly, Adaboost strategy is utilized for the arrangement of the tumor. This paper highlights the use of Artificial Neural Networks in grouping the tumors into typical or astrocytoma sort of tumors in the MRI pictures of different patients. Arun Kumar Rana et.al. [9] Machine learning makes it possible for PCs to correctly perform specific tasks. Machine learning systems are capable of executing complex tasks by learning from information rather than following encoded instructions. One of the main goals of ML is to train the systems by utilizing past experiences to handle the present concern. The foremost purpose of the authors is to briefly summarize the various ML methods and applications for Internet of Things (IoT).

II. CHARACTERISTICS OR APPLICATIONS OF DIFFERENT ML TECHNIQUES

Technique	Type	Key Algorithms	Applications	Advantages	Limitations
Supervised Learning	Classification & Regression	Decision Trees, SVM, KNN, Linear Regression	Spam detection, Image recognition	High accuracy with labeled data	Requires large labeled datasets
Unsupervised Learning	Clustering & Dimensionality Reduction	K-Means, DBSCAN, PCA	Customer segmentation, Anomaly detection	No need for labeled data	Hard to validate outcomes
Semi-supervised Learning	Mix of both	Label propagation, Graph-based models	Text classification, Medical diagnosis	Reduces labeling cost	Still depends on some labeled data
Reinforcement Learning	Decision-making	Q-Learning, Deep Q-Networks (DQN)	Robotics, Game AI, Autonomous vehicles	Learns optimal policy over time	Needs many interactions (slow learning)
Deep Learning	Neural Network-Based	CNNs, RNNs, Transformers	Speech recognition, NLP, Computer vision	Handles complex unstructured data	High computational cost

III. MACHINE LEARNING PARADIGMS

Machine Learning is an exceptional part of artificial intelligence which secures intelligence based on known absolutes from preparing data [10]. Before applying it to the real information set, machine learning approaches operate by training an algorithm with the training data set. Machine learning methods initially train the algorithm with certain specific inputs with known outputs so it can later work with new obscure data [11].

A. Supervised Learning

Supervised learning is a learning technique that reveals both the sources of information and yields. Based on this information about the training, the equation must be summarized with the end goal of being able to respond efficiently to every single feedback imaginable. This supervised calculation is used to establish correct yield for inputs not spotted during training. The supervised learning algorithms begin with gathering the training data and

then pre-processing it. After feature representation of functions, it determines the appropriate algorithm and executes it. The performance levels are measured on trained and tested data to achieve accuracy. To deal with a given problem using supervised learning measurement, some unique developments need to be sought.

- Determine the type of models to be planned.
- Collect and coordinate a collection of plans.
- Determine the details that illustrate the informed work representation.
- Determine the learning ability structure and evaluate the measurement of training.
- Complete the program and use the knowledge assembly to measure the reading.

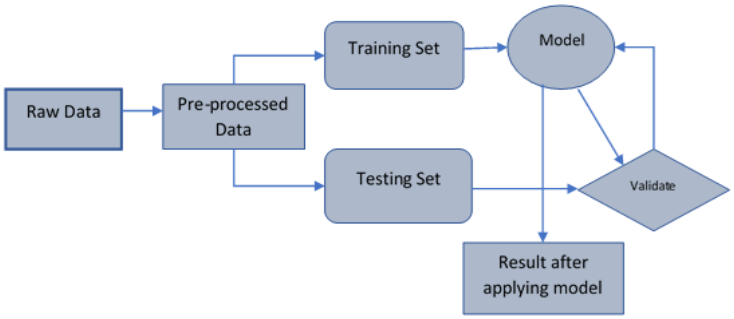

Fig. 3. Supervised Learning Workflow

B. Unsupervised Learning

Unsupervised training seems to be much harder as the aim of this technique is to get the machine to find out how to do something, we don't disclose how to do it. The principal approach is to explain the representative agent by providing some rewards rather than explicit categorizations. This method is mostly applied in decision problem frameworks as it helps to make maximum profitable decisions based on given data [12].

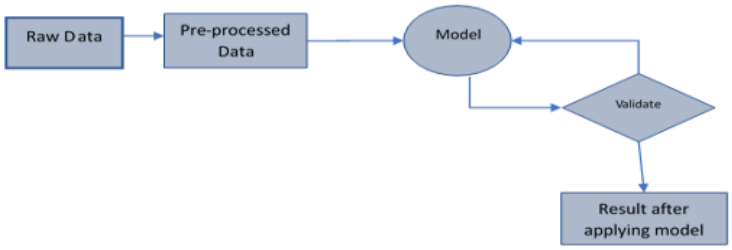

Fig.4. Unsupervised Learning Workflow

C. Semi-Supervised Learning

Semi-Supervised learning is a technique that consolidates the strength of both the above-mentioned learning techniques. A part of data labeled when the acquisition of data is accomplished by human experts is known as Semi-Supervised learning.

- Models of generation
- Low-density separation
- Graph-based methods

- Co-training [13]

D. Reinforcement Learning

In Reinforcement learning, certain goals are set and achieved by using computer interaction with an appropriate environment. A strengthening method may allow a customer (e.g., domain expert) to identify an instance that may result from multiple unlabeled instances [14].

E. Ensemble Learning

At the point when different individual students are consolidated to shape just a single student then that specific kind of learning is called ensemble learning. These individual learners can be Decision Tree, Neural Network, Naïve Bayes, and so on. It has been identified that an assortment of students is quite often better at making a specific showing as opposed to singular students [15]. Regardless of the way that it is comprehended that the shortcomings of the classifiers are assembled in the ensemble classifier, yet some blend it has delivered a skilled exhibition. Specialists were increasingly involved in classifying ensemble classifiers.

F. Multitask Learning

Multitask learning has a fundamental objective to help and support other learners for better performance. When the algorithms employed for multitask learning are applied on an undertaking, it recalls the system of how it tackled the issue or how it scopes to the specific end [16].

IV. MACHINE LEARNING ALGORITHMS

Nowadays, there is a vast interaction of optimization and machine learning for solving problems. It offers good performance in practice in terms of execution times and memory requirements. The use of these algorithms is highly done due to the usage of simple and easy implementation methods. They also have a good fast divergence to an approximate solution of design and have robustness and geometric stability for the class of machine learning models. The given below summarizes various intelligent machine algorithms and their usage in real-life applications.

A. Naive Bayes classifier

It is a supervised depiction strategy generated using Bayes with a Naive prohibitive probability hypothesis' assumption that each highlighted pair is typically autonomous. In other words, the proximity of an element is not determined by the proximity of another by any means. Because of this over-streamlined assumption, in various common-sense situations, such as content order and spam detection, NB classifiers performed very well. Just a small amount of information planning needs to evaluate those parameters. Also, besides even exceptionally propelled structure approaches are largely outlined by NB classifiers. Based on the class mark given to Naive Bayes, it is agreed that the characteristics are restrictively free and thus attempts to assess the likelihood of class-dependence.

B. Decision Tree

The role of Decision Tree (DT) is to make a classifier based on a few certainly known instances to predict the estimation of an objective class for an inconspicuous test case. A Decision Tree is arranging a discreet test

example through a collection of decisions. Decision Tree (DT) is used to predict the output of an unseen class, that is based on previously known instances, through a sequence of decisions. It can be expanded into two types:

- Classification tree, with a major class mark sprint.
- Regression tree with a numbered array of category marks.

C. K-Nearest Neighbor

The K-Nearest Neighbor (K-NN) approach can be used in numerous distance measuring techniques. K-nearest neighbor finds out k number of samples in training data that are nearest to the test sample and then it assigns the most frequent class label among the considered training samples to the test sample. It is a modest and non-parametric approach to classify the samples.

D. Artificial Neural Network

Artificial Neural Network (ANN) is a framework-care knowledge carried out by the utility of human minds. Neural networks are frequently organized in layers consisting of different interconnected centers with an incentive component. The input layer is used to feed patterns, which communicates with all the hidden layers for further processing. These hidden layers are linked together to an output layer for generating the prediction results. Examples are displayed through the information layer to the network, which transmits to at least one shielded layer where the actual preparation is completed by means of an arrangement of measured associations. At that point, the hidden layers connect to a yield layer to yield the discovery result.

E. Support Vector Machine

In the mid-1990s, the Support Vector Machine (SVM) was implemented. Fundamentally, the concept behind SVM's interference classification is to use the arrangement of information as an approximation of the common class of items or viewed as non-attack in the interference role context and to think of the rest as anomalies in this way [17]. The classifier generated by the hypothesis of supporting vector machines divides the data space into a restricted area where the normal items are excluded and the rest of the space is assumed to contain the irregularities. Support Vector Machine is a two-class classification model.

V. APPLICATIONS

Machine Learning has accomplished huge development in various real-life applications. It should be remembered that there were no substantial commercial applications of ML algorithms until 1985. Some applications of ML are described below.

A. Automation Control

ML methods are used to a great extent in software and mechanical systems. Consider, for example, using ML to learn control strategies for helicopter balanced flight and maneuvers. Google's driving vehicles use ML to prepare data from the territory collected.

B. Realistic Experimentations

Various researchers are using ML techniques for data-intensive applications in the province of science such as neuroscience, genetics implementation, astronomy, and psychology. It is also advantageous to apply these

methods for small scale applications like stock market prediction, spam filtering, weather forecasting, market surveys, etc.

C. Bio-surveillance

A few government activities used ML algorithms to track possible ailment outbreaks. Consider the West Pennsylvania RODS project that collects the verification documents in emergency rooms and the ML programming system is designed using patient profiles to identify deviant indications, their examples, and areal appropriation. Research is progressing to fuse some extra information in the framework, like over-the-counter prescriptions' buy history to give all the more preparing information. The unpredictability of this kind of intricate and dynamic informational collection can be taken care of proficiently utilizing computerized learning strategies as they were.

D. Speech Recognition

Machine Learning approaches are generally used in existing speech recognition systems as they can better train the system to attain higher accuracy. By and by, a large portion of such systems carry out training in two unmistakable stages: autonomous preparation of pre-shipping speakers and subordinate preparation of speakers after shipping.

E. Computer Vision

Of better accuracy, a larger part of late vision systems, e.g. facial recognition technology, frameworks designed for programmed characterization minute cell images, using approaches to machine learning. For example, the U.S. The Post Office uses a computer vision system with a handwriting analyzer to sort letters with visually written areas with a precision of up to 85% [19].

VI. CASE STUDY: PREDICTING DIABETES USING MACHINE LEARNING

A practical application of supervised learning is seen in healthcare, where machine learning models help predict diabetes risk using patient data. Using the Pima Indians Diabetes dataset, algorithms such as Logistic Regression, Decision Tree, and Random Forest were applied. Among them, Random Forest achieved the highest accuracy (81.5%), demonstrating its effectiveness in binary classification tasks. Features like glucose level, BMI, and age played key roles in prediction. This case highlights how machine learning enhances diagnostic accuracy, aids early intervention, and supports clinical decision-making, showcasing its transformative potential in real-world, high-impact domains like healthcare.

VII. CONCLUSION

This survey highlights the core machine learning techniques—supervised, unsupervised, semi-supervised, and reinforcement learning—and their diverse applications across domains. Current trends point toward the growing use of deep learning and hybrid models for complex tasks. However, challenges such as data quality, model interpretability, and computational demands persist. Ethical concerns and bias in decision-making also remain critical issues. Future research should focus on explainable AI, efficient learning from limited data, and domain-specific model optimization. Advancing these areas will help make machine learning more robust, transparent, and applicable to real-world problems across science, healthcare, finance, and beyond.

Author contributions

Amritpreet Kaur: Concept, Writing, Visualization

Komal Sharma: Writing, Reference

Funding: Not Funding

Acknowledgments: NA

Conflict of interest: There is no conflict of Interest

REFERENCES

[1] Saranya, K., & Jayanthy, S. (2017). Onto-based sentiment classification using machine learning techniques. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). DOI:10.1109/iciiecs.2017.8276047

[2] Iqbal Muhammad and Zhu Yan, "Supervised machine learning approaches: a survey", ICTACT Journal on Soft Computing, 5(3). DOI: 10.21917/ijsc.2015.0133.

[3] Hopkins, E. (2022). Machine learning tools, algorithms, and techniques. *Journal of Self-Governance and Management Economics*, *10*(1), 43-55.

[4] Raouf Boutaba, Mohammad A. Salahuddin, Noura Limam, Sara Ayoubi, Nashid Shahriar, Felipe Estrada-Solano and Oscar M. Caicedo, "A Comprehensive Survey on Machine Learning for Networking: Evolution, Applications, and Research Opportunities", research gate 12 May 2018.

[5] Danijel Kučak, Vedran Juričića, and Goran Đambić, "Machine learning in education - a survey of current research trends", DOI: 10.2507/29th.daaam.proceedings.059 2018.

[6] S. Sarojini Devi and Y. Radhika, "A Survey on Machine Learning and Statistical Techniques in Bankruptcy Prediction", International Journal of Machine Learning and Computing, Vol. 8, No. 2, April 2018.

[7] T. Sree Kala and Dr.A. Christy, "A Survey and Analysis of Machine Learning Algorithms for Intrusion Detection System", Journal of Adv Research in Dynamical & Control Systems, 04-Special Issue, June 2017.

[8] Astina Minz, "MR Image classification using AdaBoost for brain tumor type", 2017 IEEE 7th International Advance Computing Conference (IACC) 2017.

[9] Arun Kumar Rana, Ayodeji Olalekan Salau, Swati Gupta and Sandeep Arora, "A Survey of Machine Learning Methods for IoT and their Future Applications", Amity Journal of Computational Sciences (AJCS) Volume 2 Issue 2 ISSN: 2456-6616 2016.

[10] Rodrigues, R., Camilo-Junior, C. G., & Rosa, T. (2018). A Taxonomy for Sentiment Analysis Field. International Journal of Web Information Systems, 00-00. DOI:10.1108/ijwis-07-2017-0048

[11] Devika, M., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. Procedia Computer Science, 87, 44-49. DOI: 10.1016/j.procs.2016.05.124

[12] Ajinkya Kunjir, Basil Shaikh, "A Survey on Machine Learning Algorithms for Building Smart Systems", IJIRCCE DOI: 10.15680/IJIRCCE.2017. 0501057.

[13] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.

[14] Sumit Das, Aritra Dey, Akash Pal, and Nabamita Roy, "Applications of Artificial Intelligence in Machine Learning: Review and Prospect", International Journal of Computer Applications (0975 – 8887) Volume 115 – No. 9, April 2015.

[15] D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical *Study*", Journal of Artificial Intelligence Research, 11, Pages 169- 198, 1999.

[16] R. Caruana, "*Multitask Learning*", Machine Learning, 28, 41-75, Kluwer Academic Publishers, 1997

[17] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.

[18] Kajaree Das and Rabi Narayan Behera, “Survey on Machine Learning: Concept, Algorithms, and Applications”, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Website: www.ijircce.com Vol. 5, Issue 2, February 2017.